IPN Progress Report 42-233 • May 15, 2023

Optimizing the Photon Information Efficiency of Point-to-Point Communication Using Entanglement

Matthew Thill,* Sam Dolinar,* and Dariush Divsalar*

ABSTRACT. — In order to optimize communication over a quantum channel for practical scenarios, we must deliver demanded classical and quantum resources with as little power as possible. We consider the problem of maximizing the photon information efficiency of free space optical communication over a single mode channel subject to time-varying demands for classical information. By drawing from results on the joint capacity region for transmitting classical information (bits), quantum information (qubits), and entanglement (ebits) over the lossy bosonic channel, we are able to formulate the problem as a discrete optimization. This enables us to explicitly compute the maximum photon information efficiency for a given demand profile, and the associated time-varying average photon numbers and rates of entanglement generation and consumption necessary to achieve it. Along the way, we derive a closed-form expression for the minimal average photon number necessary to achieve a particular rate pair of bit and ebit communication in the case of zero qubit communication. Our results yield examples where we can optimize our photon information efficiency by generating entanglement during times of low demand for classical information, and later use it to communicate bits at a lower average photon number during times of higher demand.

I. Introduction

Due to a growing interest in quantum applications to security, computation, and sensing, a network of quantum channels will be necessary in the future in order to communicate both classical and quantum information, as well as entanglement. A key

---

*Communications Architectures and Research Section.

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

component to this network will involve transmitting photons over free-space optical links. A standard of performance for free space optical communication is the photon information efficiency (PIE), or how many classical bits are communicated per photon in each optical mode. When communicating over an optical channel, the PIE is given by $\frac{C}{N_S}$, where C is the classical Shannon channel capacity [1] and N_S is the average photon number of the ensemble of bosonic states used in communication. The PIE is theoretically unbounded for free-space communication, mathematically modeled as a pure-loss bosonic channel. There is, however, a fundamental tradeoff between PIE and dimensional efficiency (DIE), which is the number of bits transmitted per temporal, spatial, or polarization dimension. At low DIE, the PIE can be approached by schemes such as pulse position modulation (PPM) and photon counting detection [2]. In general, it was shown in Giovannetti et al. [3] that for a fixed constraint on the average photon number N_S, the capacity of the pure-loss bosonic channel is given by $C = g(\eta N_S)$, where η is the transmissivity of the channel (the fraction of photons which are successfully transmitted) and $g(x) := (x+1)\log_2(x+1) - x\log_2 x$, the entropy of the Bose-Einstein distribution. This capacity can be achieved by an ensemble of coherent states, which consequently maximizes the PIE at this photon number.

Since the optical channel is fundamentally quantum in nature, it can be used to transmit resources other than just classical information, including quantum information (measured in qubits) and entanglement (measured in ebits). Each of these resources has its own rate at which it can be transmitted. Holevo-Schumacher-Westmoreland (HSW) coding [4, 5] achieves the Holevo information as a rate of classical communication, and the coherent information of a quantum channel is achievable as a rate of communication for quantum information [6–8]. Some protocols simultaneously encode several of these resources in the same quantum state, which has been shown to outperform rates achieved by time-sharing between optimal schemes to transmit the resources individually [9]. Others, notably superdense coding and quantum teleportation, use entanglement as a resource to transmit classical or quantum information at a higher rate [10–13]. Together, classical information, quantum information, and entanglement share a joint capacity region which depends on both N_S and η [14–16].

In realistic settings, such as deep space communication with spacecraft, the demand for classical information from sender to receiver varies over time, and entanglement will not be an unlimited resource with which to boost classical or quantum communication. As a result, it may make sense to vary our average photon number over time in order to achieve the maximum overall PIE. Furthermore, in times of low demand for classical information, it could be optimal to use extra photon resources to generate shared entanglement between sender and receiver, which may be used to boost the classical capacity at a lower photon number in times of higher demand. In this article, we explore this problem assuming a constant channel transmissivity. In Sections II and III, we construct an optimization that allows us to compute the maximum PIE subject to a given time-varying demand for classical information, and finite stored entanglement. We simplify our problem to the case of a finite number of equal-duration time intervals in which we have constant rates of resource communication, and focus only on transmitting classical information and consuming (or generating) entanglement to maximize PIE. In Section IV, we derive a function for the minimal photon number needed to achieve given rates of bit-communication and ebit-generation/consumption. In Sections V, VI, and VII, we explicitly compute the achievable photon information efficiency for demand profiles in up to three time intervals. Using the aforementioned function, we inductively construct an expression for the minimal photon number to communicate a demand profile given an initial amount of stored entanglement. We show that even for a small number of time intervals we observe cases with substantial percent increases in PIE by dynamically storing and consuming entanglement compared to communicating at a constant rate of entanglement consumption.

II. The Entanglement Battery Problem

We consider communicating over a single-mode lossy bosonic channel. If we mathematically denote the annihilation operator of our input signal by $\hat{a}_{\text{in}}$, and that of the environment by $\hat{e}_{\text{in}}$, then the channel transforms these operators as

$$\begin{aligned} \hat{a}_{\text{out}} &= \sqrt{\eta}\ \hat{a}_{\text{in}} + \sqrt{1-\eta}\ \hat{e}_{in} \\ \hat{e}_{\text{out}} &= -\sqrt{1-\eta}\ \hat{a}_{\text{in}} + \sqrt{\eta}\ \hat{e}_{\text{in}}, \end{aligned} \tag{1}$$

where η is the transmissivity of the channel. The environment mode may be assumed to start out in a vacuum state (the pure loss channel) or a thermal state (the thermal noise channel). The output signal state is determined by tracing out the environment. Multi-mode communication can be modeled by extending these maps bilinearly to the Kronecker product of the individual mode annihilation operators.

For now, we will consider the pure-loss single mode channel, which models the case of free-space communication with noise dominated by photon loss. The thermal noise channel is less straightforward, and in which even the classical information capacity C is only known up to a minimum output entropy conjecture [3, 17]. We will work over regimes in which the transmissivity of the channel, η, can be assumed to be constant and greater than 1/2. In that scenario, Wilde et al. [16] proved that if the sender is limited to communication with an average of N_S photons per channel use, there is a tradeoff between the rates of classical communication C in bits, quantum communication Q in qubits, and entanglement generation E in ebits per channel use. This is given by the union of the regions

$$\begin{aligned} C + 2Q &\leq g(\lambda N_S) + g(\eta N_S) - g((1-\eta)\lambda N_S) && =: f_{C2Q}(\lambda, N_S), \\ Q + E &\leq g(\eta\lambda N_S) - g((1-\eta)\lambda N_S) && =: f_{QE}(\lambda, N_S)\ , \\ C + Q + E &\leq g(\eta N_S) - g((1-\eta)\lambda N_S) && =: f_{CQE}(\lambda, N_S), \end{aligned} \tag{2}$$

where $g(x) := (x+1)\log_2(x+1) - x\log_2 x$ and $\lambda \in [0,1]$. We will use the functions $f_{C2Q}(\lambda, N_S)$, $f_{QE}(\lambda, N_S)$, and $f_{CQE}(\lambda, N_S)$ to refer to the right sides of these three expressions, as written above. For a given lambda, each one of these regions is achieved by the coding ensemble $\{p_\alpha, D^{A'}(\alpha)|\psi_{TMS}\rangle\}$, where

$$|\psi_{TMS}\rangle_{AA'} := \sum_{n=0}^{\infty} \sqrt{\frac{(\lambda N_S)^n}{(\lambda N_S + 1)^{n+1}}} |n\rangle_A |n\rangle_{A'} \tag{3}$$

is a two-mode squeezed state,

$$D^{A'}(\alpha) := \exp(\alpha \hat{a}'^{\dagger} - \alpha^* \hat{a}') \tag{4}$$

is the displacement operator acting on A', and

$$p_\alpha = \frac{1}{\pi(1-\lambda)N_S} \exp\left(\frac{-|\alpha|^2}{(1-\lambda)N_S}\right) \tag{5}$$

is a Gaussian probability distribution. Alice transmits the A' system to Bob over the lossy bosonic channel $\mathcal{N}^{A'\to B}$, maintaining the A subsystem. In doing so, Alice can communicate bits or qubits to Bob, and the two can establish shared ebits of entanglement or use them in their protocols. Positive values of C, Q, and E in Equation (2) correspond to generation of resources, whereas negative values correspond to consumption. For instance, the standard quantum teleportation protocol requires Alice to use two noiseless classical bit transmissions and consume one shared ebit with Bob in order to communicate a single qubit, leading to $C = -2$, $Q = 1$, and $E = -1$. Another example is the superdense coding protocol, in which Alice uses a single qubit transmission and consumes one ebit shared with Bob to communicate two classical bits, destroying the ebit and the transmitted qubit in the process, yielding $C = 2$, $Q = -1$, and $E = -1$. As these examples help to clarify, C and Q are actually better interpreted as the transmission of a bit or a qubit noiselessly over a channel. Negative values of C and Q correspond to Alice using bit or qubit transmissions to facilitate a protocol, while positive values represent actual information transmitted to Bob. In teleportation and superdense coding, the positive values of Q and C, respectively, correspond to *simulated* uses of an actual qubit or bit channel. This idea is made concrete by Wilde [18].

For a fixed average photon number, the inequalities in (2) yield the joint rate region of (C, Q, E)-tuples for the lossy bosonic channel. The tradeoff-coding protocol described in Wilde et al. [16, 19] which achieves this region can provably outperform methods such as time-sharing in such problems as minimizing the rate of ebit consumption for a given rate of classical communication. It does not, however, account for issues such as limited entanglement or ebit storage constraints.

In this paper, we will show how to apply the tradeoff-coding region to a scenario in which we have finite entanglement and varying communication demands over time. Our intuition is that if we consume fewer ebits during times of low information demand, we can store entanglement to assist communication at times of high demand.

In doing so, we should be able to communicate at a higher PIE. We will focus on the case of zero net qubit communication ($Q = 0$) and form an optimization problem to maximize PIE, subject to a set of time-varying demands for classical information and constraints on stored entanglement. This is similar to a setup for a power network, where stored entanglement takes the place of stored energy as a limited resource. As such, we will sometimes refer to our setup as an "entanglement battery" problem.

We mention the slight caveat that the (C, Q, E) rate tuples are *net* rates of resource generation or consumption. This means, for instance, that even if $E = 0$, a protocol might at various times make use of entanglement, though the total amounts consumed and generated will balance out. Thus, as we formulate our framework, we will place constraints on the net rates of resource use over various time intervals.

III. Formulating a Discrete Optimization Problem

To concretely formulate our problem, we consider a discrete setting in which we have equal-length time intervals indexed by $t = 1, 2, ...$ During each time interval t, Alice uses an average photon number N_t to communicate to Bob at average rates (C_t, Q_t, E_t) of bits, qubits, and ebits. Note that while this discretization is intended to approximate the continuous-time communication scenario, we must assume that each interval is long enough to achieve these average rates with a physically implementable protocol. We will associate a cost of communication which is proportional to N_t, reasoning that the average photon number should determine the power consumed over interval t since our time intervals are equal length. For the time being, we will restrict our attention to the case where $Q_t = 0$ for all t. In particular, we assume that Alice is only interested in communicating classical bits to Bob. As we mentioned before, Alice may transmit qubits in the process in order to facilitate her communication protocol, but the net rate of quantum communication over the interval will be zero.

Our goal will be to optimize PIE, measured in bits per photon. To this end, we will use the notation $C(N_t, E_t)$ to denote the maximum rate at which we can communicate classical bits as a function of photon number N_t and rate of entanglement use E_t. To make our problem more concrete still, let us assume that at time t, there is a demand D_t for a rate of classical communication. We are interested in the case where $D_t \geq 0$ for all t, since Alice's goal is ultimately to transmit classical information to Bob. But in this regime, if at some time there is a lower demand for classical information, Alice and Bob could potentially use photons to establish new entanglement to assist in classical communication later. We will assume that the two share some initial entanglement allotment, $E_0 \geq 0$.

We now make several observations: First of all, for a fixed number of time instants M, assuming equal-length time intervals, the overall photon information efficiency is proportional to $\frac{\sum_{t=1}^{M} D_t}{\sum_{t=1}^{M} N_t}$. Since the demands D_t are fixed, maximizing PIE is equivalent to minimizing the sum of the average photon numbers. We have several constraints.

First, we need to make sure that the values of N_t and E_t at each time interval are sufficient to meet our demands, namely $C(N_t, E_t) \geq D_t$. We expect $C(N_t, E_t)$ to be increasing in N_t, and since our goal is to minimize $\sum_{t=1}^{M} N_t$ it seems reasonable to replace this constraint with an equality, but we will hold off on this for now. We also need to ensure that we do not exceed our allotment of stored entanglement at any time n, which is captured by the set of inequalities $E_0 + \sum_{t=1}^{n} E_t \geq 0$, for all $n \geq 1$. This allows us to write the following optimization problem for maximizing PIE:

$$
\begin{aligned}
\underset{\{N_t, E_t\}_{t=1}^{M}}{\text{minimize}} \quad & \sum_{t=1}^{M} N_t \\
\text{subject to} \quad & N_t \geq 0,\ t = 1, ..., M, \\
& C(N_t, E_t) \geq D_t,\ t = 1, ..., M, \\
& E_0 + \sum_{t=1}^{n} E_t \geq 0,\ n = 1, ..., M.
\end{aligned}
\tag{6}
$$

Seeking to translate this optimization into the context of the lossy bosonic channel, we refer to the inequalities defined in Equation (2), which project to a region in (C, E) space when we assume that $Q = 0$. We will assume that $\eta > 1/2$. Then, noting that $g(x)$ is nonnegative and increasing for $x \geq 0$, we can verify that $f_{C2Q}(\lambda, N_S)$, $f_{QE}(\lambda, N_S)$, and $f_{CQE}(\lambda, N_S)$ are all nonnegative. Since we are interested in the case of nonnegative classical communication ($C \geq 0$), for a fixed λ and N_S we have the following necessary upper bound on the entanglement generated:

$$
\begin{aligned}
E &\leq \min(f_{QE}(\lambda, N_S),\ f_{CQE}(\lambda, N_S)) && (7) \\
&= f_{QE}(\lambda, N_S), && (8)
\end{aligned}
$$

where the equality follows from the fact that $g(x)$ is an increasing function.

For given values of λ, N_S, and $E \leq f_{QE}(\lambda, N_S)$, we have that rate of bit transmission C is bounded as

$$
\begin{aligned}
C &\leq \min(f_{C2Q}(\lambda, N_S),\ f_{CQE}(\lambda, N_S) - E) && (9) \\
&= \min(f_{CQE}(\lambda, N_S) + g(\lambda N_S),\ f_{CQE}(\lambda, N_S) - E) && (10) \\
&= f_{CQE}(\lambda, N_S) + \min(g(\lambda N_S),\ -E). && (11)
\end{aligned}
$$

It follows that if $-g(\lambda N_S) \leq E \leq f_{QE}(\lambda, N_S)$, then $C \leq f_{CQE}(\lambda, N_S) - E$. Furthermore, if $E < -g(\lambda N_S)$, indicating a larger amount of entanglement *consumption*, then there is no gain in the achievable rate of classical communication, which is upper bounded by $f_{CQE}(\lambda, N_S) + g(\lambda N_S) < f_{CQE}(\lambda, N_S) - E$. For the sake of adapting the optimization from Equation (6) to the lossy bosonic channel, this indicates that we need only perform the optimization over entanglement values E_t between $-g(\lambda_t N_t)$ and $f_{QE}(\lambda_t, N_t)$, where λ_t is the particular value of λ used in time interval t. Since the upper bound on capacity is achievable, then for a fixed value of λ_t we have $C(N_t, E_t) = f_{CQE}(\lambda_t, N_t) - E_t$ for $-g(\lambda_t N_t) \leq E_t \leq f_{QE}(\lambda_t, N_t)$.

Combining these observations, and noting the forms of the functions $f_{CQE}(\lambda, N_S)$ and $f_{QE}(\lambda, N_S)$, we can rewrite the optimization in Equation (6) in the form

$$\begin{aligned} \underset{\{\lambda_t, N_t, E_t\}}{\text{minimize}} \quad & \sum_{t=1}^{M} N_t \\ \text{subject to} \quad & g(\eta N_t) - g((1-\eta)\lambda_t N_t) \geq D_t + E_t, \\ & -g(\lambda_t N_t) \leq E_t \leq g(\eta \lambda_t N_t) - g((1-\eta)\lambda_t N_t), \\ & E_0 + \sum_{t=1}^{n} E_t \geq 0,\ n = 1, ..., M, \\ & \lambda_t \in [0,1]. \end{aligned} \tag{12}$$

We return now to the question of whether we can set $C(N_t, E_t)$ equal to D_t in the optimization from Equation (6) without loss of generality, which in Equation (12) amounts to replacing the constraint $g(\eta N_t) - g((1-\eta)\lambda_t N_t) \geq D_t + E_t$ with an equality. We argue that this is an equivalent optimization. To see this, note that we can replace the first two constraints in Equation (12) with the single expression:

$$-g(\lambda_t N_t) \leq E_t \leq \min\{f_1(\lambda_t, N_t),\ f_2(\lambda_t, N_t)\}, \tag{13}$$

where

$$f_1(\lambda_t, N_t) := g(\eta \lambda_t N_t) - g((1-\eta)\lambda_t N_t), \tag{14}$$

$$f_2(\lambda_t, N_t) := g(\eta N_t) - D_t - g((1-\eta)\lambda_t N_t). \tag{15}$$

Now suppose we have a feasible set of values $\{\lambda_t, N_t, E_t\}_{t=1}^{M}$. If E_t is strictly less than its upper bound, we can increase it and produce a new feasible point which has the same value of $\sum_{t=1}^{M} N_t$. Thus, assume without loss of generality that $E_t = \min\{f_1(\lambda_t, N_t),\ f_2(\lambda_t, N_t)\}$. If $f_2(\lambda_t, N_t) \leq f_1(\lambda_t, N_t)$, we have the desired equality. Otherwise, since $-g(\lambda_t N_t) \leq 0$ and $f_1(\lambda_t, N_t)$ is nonnegative, we are in the situation

$$-g(\lambda_t N_t) \leq 0 \leq E_t = f_1(\lambda_t, N_t) < f_2(\lambda_t, N_t). \tag{16}$$

Note that $f_1(\lambda_t, N_t)$ is increasing in λ_t, while $f_2(\lambda_t, N_t)$ is decreasing. Furthermore, when $\lambda_t = 1$, we clearly have $f_2(\lambda_t, N_t) \leq f_1(\lambda_t, N_t)$. Thus, we can increase λ_t until $0 \leq E_t \leq f_1(\lambda_t, N_t) = f_2(\lambda_t, N_t)$, and again increase E_t to to be equal to both functions. We end up with a new feasible point with the same value of $\sum_{t=1}^{M} N_t$ such that $g(\eta N_t) - g((1-\eta)\lambda_t N_t) = D_t + E_t$ as desired. Our optimization now becomes

$$\begin{aligned} \underset{\{\lambda_t, N_t, E_t\}}{\text{minimize}} \quad & \sum_{t=1}^{M} N_t \\ \text{subject to} \quad & g(\eta N_t) - g((1-\eta)\lambda_t N_t) = D_t + E_t, \\ & -g(\lambda_t N_t) \leq E_t \leq g(\eta \lambda_t N_t) - g((1-\eta)\lambda_t N_t), \\ & E_0 + \sum_{t=1}^{n} E_t \geq 0,\ n = 1, ..., M, \\ & \lambda_t \in [0,1]. \end{aligned} \tag{17}$$

The optimization in Equation (17) may seem aggravatingly similar to that in Equation (12), but as we will see shortly, introducing the equality constraint enables us to find a closed form expression for the optimal photon number at a given time t, in terms of the corresponding values of D_t and E_t. For convenience, we will define the following notation for the minimal photon number sum in the above optimization:

Definition 1. Let $S_M^{\text{opt}}(D_1, ..., D_M, E_0)$ be the minimal value of $\sum_{t=1}^M N_t$ achieved in the optimization in Equation (17) for M time intervals, initial shared entanglement E_0, and classical information demands $D_1, ..., D_M$.

IV. Optimal Photon Number

One difficulty with the optimization in Equation (17) is that the parameter space over the variables $\{\lambda_t, N_t, E_t\}_{t=1}^M$ is large and high-dimensional. To make the problem more tractable, it is desirable to find ways to eliminate some of these variables. With a bit of work, it turns out that we can reduce this optimization to one which is only in terms of $\{E_t\}_{t=1}^M$. The first step toward this direction is to observe that we can compute the optimal photon number N_t at time t as a function of the pair (D_t, E_t). The following theorem describes this function, and gives a simple description of the set of feasible E_t. First, we must define three functions in terms of $g(x)$ and η:

$$h(x) := g(\eta x) - g((1-\eta)x), \tag{18}$$

$$L(x) := g^{-1}(|x|) \cdot \mathbf{I}(x < 0) + h^{-1}(|x|) \cdot \mathbf{I}(x \geq 0), \tag{19}$$

$$d(x) := h(g^{-1}(x)) + x. \tag{20}$$

Here, $\mathbf{I}(\cdot)$ is an indicator function. Note that $h(x)$ is increasing for $\eta > 1/2$, which is our regime of interest.

We will also make use of the following simple lemma:

Lemma 1. The function $x \log\left(1 + \frac{1}{x}\right)$ is increasing for $x > 0$.

Proof. This function has derivative $\log\left(\frac{x+1}{x}\right) - \frac{1}{x+1} = \int_x^{x+1} \frac{1}{t} dt - \frac{1}{x+1}$. The lemma follows from the fact that since $\frac{1}{t}$ is lower bounded by $\frac{1}{x+1}$ over the unit interval $[x, x+1]$, then $\int_x^{x+1} \frac{1}{t} dt \geq ((x+1) - x) \cdot \frac{1}{x+1} = \frac{1}{x+1}$.

We are now equipped to state the theorem:

Theorem 1. If $E_t \geq -d^{-1}(D_t)$, then the minimal value of N_t such that the pair (N_t, E_t) is feasible in optimization in Equation (17) is given by the expression

$$N_t^{\text{opt}}(D_t, E_t) = \frac{1}{\eta} g^{-1}\left(D_t + E_t + g((1-\eta)L(E_t))\right). \tag{21}$$

If $E_t < -d^{-1}(D_t)$, then E_t is not feasible in optimization in Equation (17).

Proof. Consider the following relaxation of optimization from Equation (17):

$$\begin{aligned}
\underset{\{L_t, N_t, E_t\}}{\text{minimize}} \quad & \sum_{t=1}^{M} N_t \\
\text{subject to} \quad & g(\eta N_t) - g((1-\eta)L_t) = D_t + E_t, \\
& -g(L_t) \leq E_t \leq g(\eta L_t) - g((1-\eta)L_t), \\
& E_0 + \sum_{t=1}^{n} E_t \geq 0, \;\; n = 1, ..., M.
\end{aligned} \tag{22}$$

Here we have replaced the constrained variable λ_t and with the unconstrained variable L_t, which is a surrogate for the product $\lambda_t N_t$. Rewriting the equality constraint as

$$g(\eta N_t) = D_t + E_t + g((1-\eta)L_t), \tag{23}$$

or equivalently

$$N_t = \frac{1}{\eta} g^{-1}(D_t + E_t + g((1-\eta)L_t)), \tag{24}$$

and noting that $g(\cdot)$ is an increasing function, we see that for fixed values of D_t and E_t the minimal N_t is achieved when L_t is minimal. L_t is constrained by the inequality relations $-g(L_t) \leq E_t \leq g(\eta L_t) - g((1-\eta)L_t)$.

If $E_t < 0$, then L_t is minimized when $-g(L_t) = E_t$, or rather $L_t = g^{-1}(-E_t) = g^{-1}(|E_t|)$. Using this value for L_t and the corresponding value for N_t from Equation (24), then if $L_t/N_t \leq 1$ we can set $\lambda_t = L_t/N_t$ to form a feasible set of values in the original optimization from Equation (17). Furthermore, since L_t is minimal, we have that the optimal value of N_t is $\frac{1}{\eta} g^{-1}(D_t + E_t + g((1-\eta)g^{-1}(|E_t|)))$. On the other hand, if $L_t/N_t > 1$ at the minimal value of L_t, then we must seek a feasible value of $\lambda_t = L_t/N_t$ at a higher value of L_t. Since we need $\lambda_t \leq 1$, we note that this is equivalent to requiring that $L_t \leq N_t$. From Equations (23) and (24), and the fact that $g(\cdot)$ is an increasing function, we can translate this to the condition that the difference

$$g(\eta L_t) - (D_t + E_t + g((1-\eta)L_t)) \tag{25}$$

be less than or equal to 0. By assumption, when $L_t = g^{-1}(-E_t)$ we have $L_t > N_t$, so this difference is positive. Differentiating Equation (25) with respect to L_t, we get

$$\begin{aligned}
&\frac{d}{dL_t}\left[g(\eta L_t) - (D_t + E_t + g((1-\eta)L_t))\right] \\
&\quad = \eta g'(\eta L_t) - (1-\eta) g'((1-\eta)L_t) \\
&\quad = \frac{1}{L_t \log 2}\left[\eta L_t \log\left(1 + \frac{1}{\eta L_t}\right) - (1-\eta) L_t \log\left(1 + \frac{1}{(1-\eta)L_t}\right)\right].
\end{aligned} \tag{26}$$

Since we assume $\eta \geq 1/2$, and $x \log\left(1 + \frac{1}{x}\right)$ is an increasing function, this derivative is nonnegative for $L_t > 0$. In other words, the difference in (25) is increasing with respect to L_t, so there is no feasible value of λ_t in optimization from Equation (17).

We can formulate the condition for a feasible λ_t more cleanly as follows: when $L_t = g^{-1}(-E_t)$ we need the difference in Equation (25) to be nonnegative. In other words,

$$g(\eta g^{-1}(-E_t)) - g((1-\eta)g^{-1}(-E_t)) - E_t \geq D_t. \tag{27}$$

Using the functions we defined in Equations (18) and (20), we can express this as the relation $d(-E_t) \geq D_t$. Since $d(\cdot)$ is a composition of increasing functions, it is also increasing, so this is equivalent to $E_t \geq -d^{-1}(D_t)$ as in the theorem statement.

Now suppose instead that $E_t \geq 0$. Then L_t is minimized by setting $g(\eta L_t) - g((1-\eta)L_t) = E_t$, so that $L_t = h^{-1}(E_t)$. We again set N_t to $\frac{1}{\eta}g^{-1}(D_t + E_t + g((1-\eta)L_t))$ to satisfy the equality constraint in Equation (22). By the same reasoning as before, if $\lambda_t := L_t/N_t \leq 1$, then λ_t is feasible in optimization from Equation (17) and this value of N_t is optimal. Also by our previous argument, if $L_t/N_t > 0$, then since L_t is minimal, there is no feasible value for λ_t. In this case, our criterion for a feasible λ_t is $L_t \leq N_t$ which corresponds to

$$h^{-1}(E_t) \leq \frac{1}{\eta}g^{-1}(D_t + E_t + g((1-\eta)h^{-1}(E_t))) \tag{28}$$

$$\iff g(\eta h^{-1}(E_t)) - g((1-\eta)h^{-1}(E_t)) - E_t \leq D_t \tag{29}$$

$$\iff h(h^{-1}(E_t)) - E_t \leq D_t \tag{30}$$

$$\iff 0 \leq D_t. \tag{31}$$

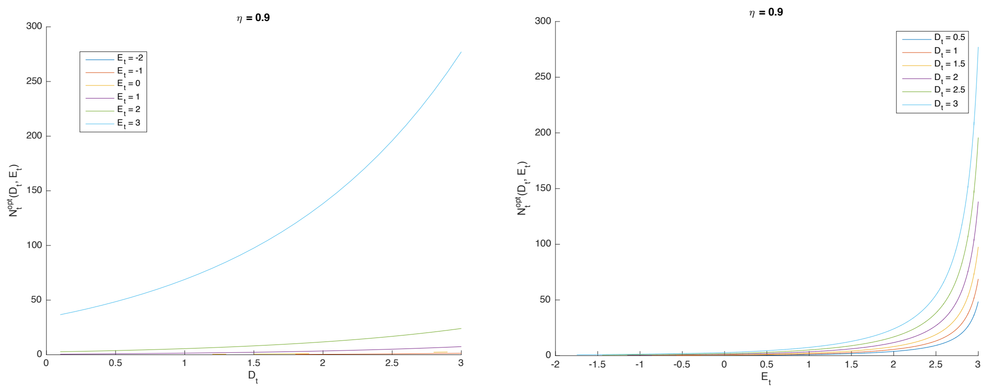
Since we assume the demands D_t are nonnegative, this condition is always met, so all nonnegative values of E_t merit feasible solutions to optimization from Equation (17). Thus, our criterion $E_t \geq -d^{-1}(D_t)$ defines all feasible pairs (D_t, E_t).

We complete the proof by noting that the function $L(\cdot)$ from Equation (19) is defined such that $L(E_t) = L_t$ for any feasible E_t, so that the expression for $N_t^{\text{opt}}(D_t, E_t)$ in Equation (21) is consistent with the optimal values of N_t we derived for both negative and nonnegative E_t.

In Figure 1, we depict the function $N_t^{\text{opt}}(D_t, E_t)$ by plotting with respect to D_t and E_t for $\eta = 0.9$. Note that in Figure 1(a), we can see that as the entanglement generation increases from $E_t = 2$ to $E_t = 3$, the optimal photon number jumps dramatically over all demands D_t plotted. In terms of our optimization in Equation (17), this could indicate a need to generate entanglement conservatively in certain regimes.

A. The Question of Convexity and Properties of $N_t^{\text{opt}}(D_t, E_t)$

Since the plots in Figure 1 suggest that the function $N_t^{\text{opt}}(D_t, E_t)$ is convex with respect to each of D_t and E_t, a natural question is whether our optimization in Equation (17) is a convex problem. This is evidently not the case, as we can see by examining the feasible region of the points (E_t, D_t) carved out by the constraint $E_t \geq -d^{-1}(D_t)$. We plot the curve $E_t = -d^{-1}(D_t)$ lower-bounding this region in



(a) **(b)**

Figure 1. The optimal photon number $N_t^{\mathrm{opt}}(D_t, E_t)$ plotted with respect to (a) the demand for classical communication, D_t, for fixed values of consumed or generated entanglement E_t, and (b) the value of E_t for fixed values of D_t. In both plots, we have fixed the transmissivity at $\eta = 0.9$.

Figure 2(a), and the detrended version of the curve (subtracting the best straight-line fit) in Figure 2(b). The feasible (E_t, D_t) region is represented by the area above these curves, and from the detrended plot, we can see that this region is not convex.

One thing we can show, however, is the following:

Lemma 2. The function $N_t^{\mathrm{opt}}(D_t, E_t)$ is increasing in both D_t and E_t for $1/2 \leq \eta \leq 1$.

Proof. First, note that if (D_t, E_t) is a feasible point (meaning that $E_t + d^{-1}(D_t) \geq 0$), then increasing either E_t or D_t will yield a new feasible point. This is because $d^{-1}(x)$ is a sum of x and $h \circ g^{-1}(x)$, a composition of increasing functions, hence $d^{-1}(x)$ is itself increasing. For a fixed E_t, $N_t^{\mathrm{opt}}(D_t, E_t)$ is increasing in D_t because $g^{-1}(\cdot)$ is an increasing function. Fixing D_t instead, we consider two regimes: $E_t \geq 0$ and $E_t < 0$.

If $E_t \geq 0$, then $N_t^{\mathrm{opt}}(D_t, E_t) = \frac{1}{\eta} g^{-1}(D_t + E_t + g((1-\eta)h^{-1}(E_t))$. Then, since $g^{-1}(\cdot)$ is increasing and $g((1-\eta)h^{-1}(\cdot))$ is a composition of increasing functions, $N_t^{\mathrm{opt}}(D_t, E_t)$ is increasing in E_t.

If, on the other hand, $E_t < 0$, we have $N_t^{\mathrm{opt}}(D_t, E_t) = \frac{1}{\eta} g^{-1}(D_t + f_g(E_t))$, where we define $f_g(x) := x + g((1-\eta)g^{-1}(-x))$. Note that since $g(x)$ is nonnegative and increasing for $x \geq 0$, we have $f_g(x) \geq x + g(g^{-1}(-x)) = 0$, so $f_g(x)$ is also nonnegative, and strictly positive when $\eta > 0$ and $x > 0$. Examining the partial derivative with respect to E_t:

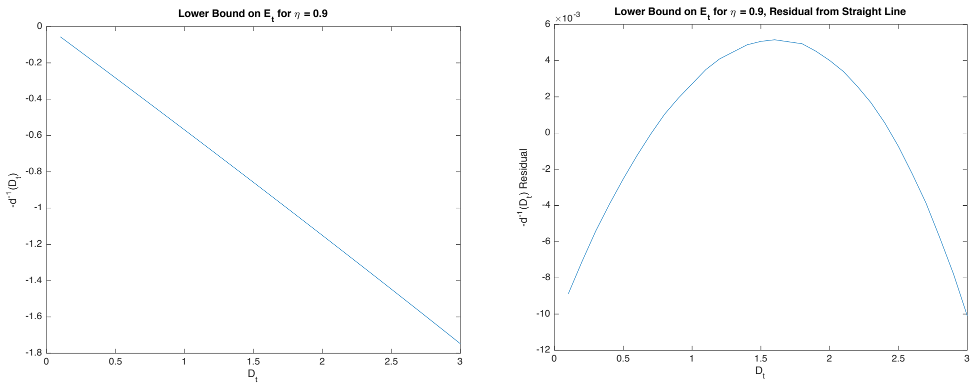
$$\frac{\partial N_t^{\mathrm{opt}}}{\partial E_t} = \frac{1}{\eta} \cdot \frac{1}{g'(g^{-1}(D_t + f_g(E_t)))} \cdot \left[1 - (1-\eta) \cdot \frac{g'((1-\eta)g^{-1}(-E_t))}{g'(g^{-1}(-E_t))}\right]. \tag{32}$$

Now, noting that $g'(x) = \log_2(1 + x^{-1})$ is positive for $x > 0$, we see that the sign of $\frac{\partial N_t^{\mathrm{opt}}}{\partial E_t}$ is completely determined by that of $\left[1 - (1-\eta) \cdot \frac{g'((1-\eta)g^{-1}(-E_t))}{g'(g^{-1}(-E_t))}\right]$.

Furthermore, since $g'(x)$ is decreasing for $x > 0$, we have that $\frac{g'((1-\eta)g^{-1}(-E_t))}{g'(g^{-1}(-E_t))} \leq 1$, hence

$$\left[1-(1-\eta)\cdot\frac{g'((1-\eta)g^{-1}(-E_t))}{g'(g^{-1}(-E_t))}\right] \geq [1-(1-\eta)] = \eta.$$

It follows that the partial derivative of $N_t^{\text{opt}}(D_t, E_t)$ with respect to E_t is positive, which concludes the proof.



(a)

(b)

Figure 2. Outline of the feasible region $E_t \geq -d^{-1}(D_t)$. In (a) we plot the lower bound on E_t for $\eta = 0.9$, and in (b) we plot the detrended version of this curve, subtracting the best straight-line fit. The feasible region lies above each of these curves. The detrended plot indicates that the region is not convex.

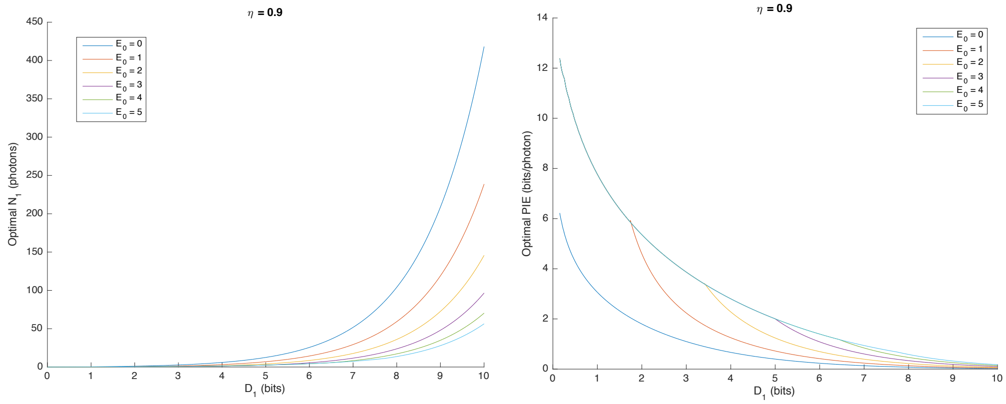
V. Solving for One Time Instant

Consider the simple case of a single time interval, $M = 1$, with a classical information demand D_1 and some initial entanglement storage E_0. More concretely, we mean that over a single interval we can afford to consume entanglement at a maximal rate of E_0. Here, it is optimal to have $E_1 \leq 0$ since any generated entanglement could not be used at later time intervals and would only increase our photon number N_1. In fact, assuming $E_0 \geq 0$ and the fact that $N_t^{\text{opt}}(D_t, E_t)$ is increasing in E_t, we see that it is optimal to choose $E_1 = \max(-E_0, -d^{-1}(D_1))$, which is less than or equal to 0. The minimal average photon number for a single time interval is then given as

$$\begin{aligned} S_1^{\text{opt}}(D_1, E_0) &= N_1^{\text{opt}}(D_1, E_1) && (33)\\ &= \frac{1}{\eta} g^{-1}\left(D_1 + E_1 + g((1-\eta)g^{-1}(|E_1|))\right), && (34)\end{aligned}$$

where $E_1 = \max(-E_0, -d^{-1}(D_1))$.

We plot the optimal average photon number and the associated optimal PIE at transmissivity $\eta = 0.9$ and varying values of $E_0 \geq 0$ in Figures 3(a) and 3(b), respectively. As we can see, for low enough D_1, stored entanglement improves the PIE, driving it up to an upper bound related to the maximum capacity gain given unlimited entanglement. But as D_1 increases, the PIE begins to drop, asymptotically approaching the PIE in the absence of shared entanglement ($E_0 = 0$).



(a)

(b)

Figure 3. (a) Optimal average photon number to meet classical communication rate D_1 at varying initial entanglement storages E_0 over a single time interval ($M = 1$). (b) The associated optimal photon information efficiency, D_1/N_1. In both plots, transmissivity is $\eta = 0.9$.

VI. Solving for Two Time Instants

Next, we consider the $M = 2$ scenario, where we again have an initial entanglement storage $E_0 \geq 0$ and now two successive demands for rates of classical communication, D_1 and D_2. We want to find associated rates of entanglement generation/consumption, E_1 and E_2, to minimize the total photon number $N_1^{\text{opt}}(D_1, E_1) + N_2^{\text{opt}}(D_2, E_2)$, where we assume that we will communicate at the optimal average photon number within each of the two time intervals. We have four constraints on E_1 and E_2. First, we must have $E_1 \geq -d^{-1}(D_1)$ and $E_2 \geq -d^{-1}(D_2)$ in order for both to be feasible by Theorem 1. The third and fourth constraints are the entanglement constraints from optimization in Equation (17), namely that $E_0 + E_1 \geq 0$ and $E_0 + E_1 + E_2 \geq 0$. As such, we will assume that the following implied constraint on E_0 is satisfied:

$$E_0 \leq d^{-1}(D_1) + d^{-1}(D_2). \tag{35}$$

Were this not the case, we would have more entanglement than we could effectively use to boost our capacity at either time instant. It would be optimal to set $E_1 = -d^{-1}(D_1)$ and $E_2 = -d^{-1}(D_2)$.

We argue that $E_2 \leq 0$, reflecting that there is no need to generate new shared entanglement in the final time interval since it cannot be used later. Stated more rigorously, if E_1 and E_2 satisfy the above constraints with E_2 positive, then setting $E_2 = 0$ would still satisfy them and would yield a lower value of $N_2^{\text{opt}}(D_2, E_2)$ by Lemma 2. Furthermore, since the optimal E_2 is less than or equal to zero, the condition $E_0 + E_1 \geq 0$ is implied by the contraint $E_0 + E_1 + E_2 \geq 0$.

Now, suppose that the optimal values of E_1 and E_2 satisfy the entanglement constraint strictly: $E_1 + E_2 > -E_0$. Then by Equation (35) we see that E_t must be strictly greater than $-d^{-1}(D_1)$ for either $t = 1$ or $t = 2$. Since $N_t^{\mathrm{opt}}(D_t, E_t)$ is increasing by Lemma 2, we can decrease the corresponding E_t while maintaining our entanglement constraint, lowering the sum $N_1^{\mathrm{opt}}(D_1, E_1) + N_2^{\mathrm{opt}}(D_2, E_2)$ and achieving a more optimal PIE. Thus, we see that, assuming Equation (35), the optimal pair (E_1, E_2) must satisfy $E_1 + E_2 = -E_0$. We may thus simplify our optimization by setting $E_1 = -E_0 - E_2$, and optimizing only over feasible values of E_2, giving the relaxed optimization

$$\begin{aligned}
\underset{\{E_2\}}{\text{minimize}} \qquad & \frac{1}{\eta} g^{-1}(D_1 - E_0 - E_2 + g((1-\eta)L(-E_0 - E_2))) \\
& \quad + \frac{1}{\eta} g^{-1}(D_2 + E_2 + g((1-\eta)g^{-1}(-E_2))) \\
\text{subject to} \qquad & E_2 \in [-d^{-1}(D_2), 0]
\end{aligned} \tag{36}$$

As a sanity check, we must confirm that the relaxed optimization in Equation (36) returns a value of $E_1 = -E_0 - E_2$ greater than or equal to $-d^{-1}(D_1)$. If not, then the condition in Equation (35) implies that it must return a value of E_2 strictly greater than $-d^{-1}(D_2)$. In this case, though, the two summands $\frac{1}{\eta} g^{-1}(D_1 + E_1 + g((1-\eta)L(E_1)))$ and $\frac{1}{\eta} g^{-1}(D_2 + E_2 + g((1-\eta)g^{-1}(-E_2)))$ of the optimized function could both be minimized by raising E_1 to $-d^{-1}(D_1)$, and simultaneously lowering E_2 to $-E_0 - E_1 \geq -d^{-1}(D_2)$.

In Figure 4, we plot the function optimized in Equation (36) with respect to E_2 varied over its feasible range. We consider demands D_1 and D_2 summing to 5, and fix our prior stored entanglement at $E_0 = 1$ and our transmissivity at $\eta = 0.9$. As we can see, examining the minima of the curves and the corresponding values of E_2 and $E_1 = 1 - E_2$, when $D_1 << D_2$, the optimal PIE is achieved when $E_1 > 0$. Intuitively this means that when the initial demand for classical information is low, it benefits us to invest photons in generating shared entanglement which can then be used to boost capacity in the second time interval. As D_1 grows with respect to D_2, there is a point beyond which it becomes optimal to generate no new entanglement during the first time interval: $E_1 \leq 0$.

There are two cases in which we can determine the optimal entanglement used at each time instance. The first is the case of equal demand in both intervals: $D_1 = D_2$. In this case, speculating the optimized function in Equation (36), we see the minimal photon number occurs when $E_1 = E_2 = -\frac{E_0}{2}$. The other case is when $\eta \approx 1$. In this case, the optimized function in Equation (36) becomes approximately $\frac{1}{\eta}\left[g^{-1}(D_1 + E_1) + g^{-1}(D_2 + E_2)\right]$. This achieves its minimum when $D_1 + E_1 \approx D_2 + E_2$, or $E_2 \approx \frac{1}{2}(D_1 - D_2 - E_0)$, if feasible.

In Figure 5, we depict the effects of varying the initial stored entanglement E_0 in the $M = 2$ optimization. Figure 5(a) shows the average total photon number $S_2^{\mathrm{opt}}(D_1, D_2, E_0)$, again fixing D_1 and D_2 to sum to 5 and setting $\eta = 0.9$, plotted

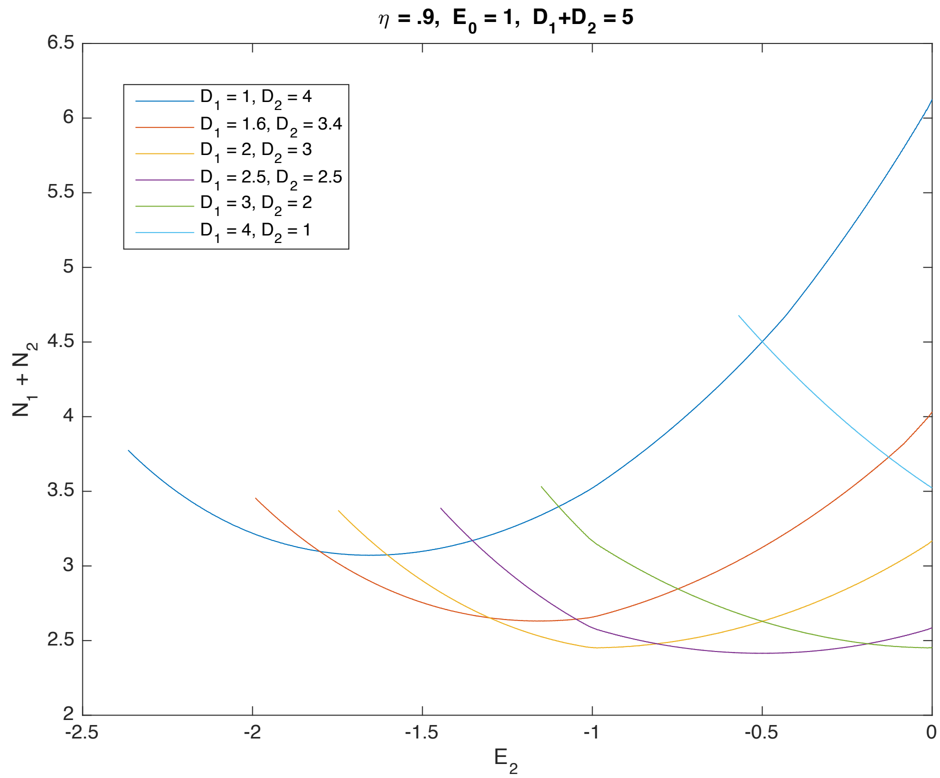


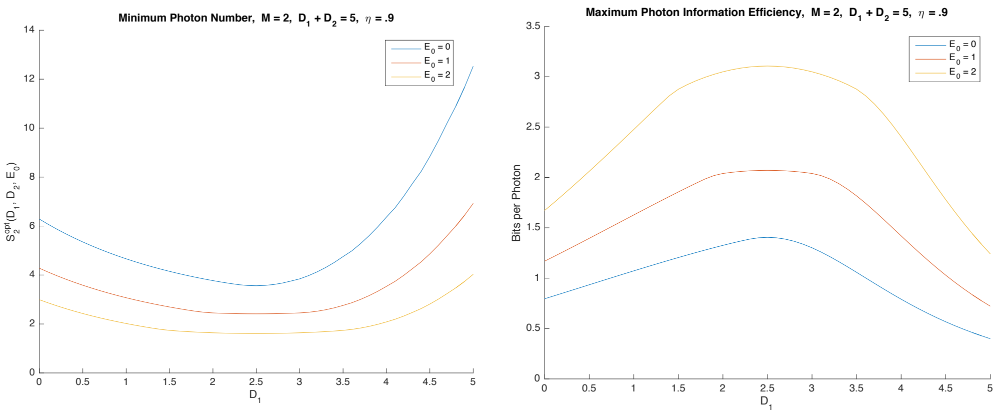
Figure 4. Optimal average photon numbers over two time intervals. Plotted are minimal values of $N_1 + N_2$ for varying demand pairs (D_1, D_2) satisfying $D_1 + D_2 = 5$. The initial entanglement storage is fixed at $E_0 = 1$ for all cases, and the transmissivity is $\eta = 0.9$.

with respect to D_1. Figure 5(b) shows the resulting average photon information efficiency, $5/S_2^{\text{opt}}(D_1, D_2, E_0)$. As we can see, raising E_0 from 0 to 2 nearly triples the number of bits communicated per photon over the two time intervals for certain values of D_1 and D_2.

VII. $M > 2$ Time Instances

Consider the more general case of M time intervals. For demands $D_1, ..., D_M$ and a given set of feasible $E_1, ..., E_M$, Theorem 1 shows that the optimal photon number at time interval t is $N_t^{\text{opt}}(D_t, E_t)$. Thus, we can reformulate our entanglement battery optimization as

$$\begin{aligned} \underset{\{E_t\}}{\text{minimize}} \quad & \sum_{t=1}^{M} N_t^{\text{opt}}(D_t, E_t) \\ \text{subject to} \quad & E_t \geq -d^{-1}(D_t),\ t = 1, ..., M, \\ & E_0 + \sum_{t=1}^{n} E_t \geq 0,\ n = 1, ..., M. \end{aligned} \tag{37}$$



(a) **(b)**

Figure 5. (a) Optimal average photon number to meet classical communication rates for $M = 2$ time instants, D_1 and D_2, which sum to $D_1 + D_2 = 5$. We plot with respect to D_1 at varying initial entanglement storages E_0, fixing transmissivity $\eta = 0.9$. (b) The associated optimal PIE.

We can generalize some of the observations we noticed in the $M = 2$ case. First, we should never generate new entanglement in the final time interval:

Lemma 3. An optimal point $(E_1, ..., E_M)$ in optimization from Equation (37) satisfies $E_M \leq 0$.

Proof. Suppose a feasible point $(E_1, ..., E_M)$ satisfies $E_M > 0$. Then the point $(E_1, ..., E_{M-1}, 0)$ is still feasible, because the condition $E_0 + \sum_{t=1}^{M-1} E_t \geq 0$ implies that $E_0 + \sum_{t=1}^{M} E_t \geq 0$ for $E_M = 0$. The new point also has a lower value of $N_M^{\text{opt}}(D_M, E_M)$ by Lemma 2, hence is a more optimal point.

We can also argue the following:

Lemma 4. If $E_0 < d^{-1}(D_1) + ... + d^{-1}(D_M)$, then an optimal point $(E_1, ..., E_M)$ in optimization from Equation (37) will satisfy $E_0 + \sum_{t=1}^{M} E_t = 0$. If $E_0 \geq d^{-1}(D_1) + ... + d^{-1}(D_M)$, then setting $E_t = -d^{-1}(D_t)$ yields the optimal point.

Proof. If $E_0 \geq \sum_{t=1}^{M} d^{-1}(D_t)$, then the set of points defined by $E_t = -d^{-1}(D_t)$, $t = 1, ..., M$, are clearly feasible, and by Lemma 2 they are optimal since each E_t takes on its minimal value.

If $E_0 < \sum_{t=1}^{M} d^{-1}(D_t)$, then consider a feasible point $(E_1, ..., E_M)$ and suppose the constraint $E_0 + \sum_{t=1}^{M} E_t \geq 0$ is a strict inequality (> 0) in optimization from Equation (37). This implies that $\sum_{t=1}^{M} E_t > \sum_{t=1}^{M} -d^{-1}(D_t)$, which means that at least one E_t is strictly greater than $-d^{-1}(D_t)$. Let t^* be the maximum such index, so that $E_t = -d^{-1}(D_t)$ for $t = t^* + 1, ..., M$. Since these last E_t are less than or equal to 0, we see that $E_0 + \sum_{t=1}^{m} E_t > 0$ for $m = t^*, t^* + 1, ..., M$. We can then lower E_{t^*} to $\max(-d^{-1}(D_{t^*}), \{-E_0 - \sum_{t=1}^{m} E_t\}_{m=t^*}^{M})$, obtaining a new point which is still feasible and has a lower value of $N_{t^*}^{\text{opt}}(D_{t^*}, E_{t^*})$ by Lemma 2, hence is more optimal. It follows that the optimal feasible point satisfies $E_0 + \sum_{t=1}^{M} E_t = 0$.

We can now solve the entanglement battery problem for higher numbers of time intervals by using an inductive argument, as described in the following theorem:

Theorem 2. Consider the entanglement battery optimization in Equation (37) for M time intervals, classical information demands D_1, ..., D_M, and initial stored entanglement E_0.

If $E_0 \geq \sum_{t=1}^{M} d^{-1}(D_t)$, then the optimal point of optimization in Equation (37) occurs when $E_t = -d^{-1}(D_t)$ for each t, and the minimal value is given by $\sum_{t=1}^{M} N_t^{\text{opt}}(D_t, -d^{-1}(D_t))$.

If $E_0 < \sum_{t=1}^{M} d^{-1}(D_t)$, then the solution can be computed inductively as

$$S_M^{\text{opt}}(D_1, ..., D_M, E_0) = \min_{E_1 \in \mathcal{I}_{E_1}} N_1^{\text{opt}}(D_1, E_1) + S_{M-1}^{\text{opt}}(D_2, ..., D_M, E_0 + E_1), \tag{38}$$

where $\mathcal{I}_{E_1} := \Big[\max(-E_0, -d^{-1}(D_1)), \Big(\sum_{t=2}^{M} d^{-1}(D_t)\Big) - E_0\Big]$.

Proof. The case $E_0 \geq \sum_{t=1}^{M} d^{-1}(D_t)$ was shown in Lemma 4, so we will focus on the case $E_0 < \sum_{t=1}^{M} d^{-1}(D_t)$. Consider a feasible point $(E_1, ..., E_M)$. After the time interval $t = 1$, the remaining stored entanglement is $E_0 + E_1$ (which is nonnegative by the constraints in Equation (37)) and there are $M - 1$ remaining time intervals with demands $D_2, ..., D_M$. Thus, given E_1, the values $E_2, ..., E_M$ are optimized by solving the entanglement battery problem over $M - 1$ time intervals, which yields optimal photon number $S_{M-1}^{\text{opt}}(D_2, ..., D_M, E_0 + E_1)$. If $E_0 + E_1 \geq \sum_{t=2}^{M} d^{-1}(D_t)$, then by Lemma 4 the optimal entanglement values will be $E_t = -d^{-1}(D_t)$ for $t = 2, ..., M$, and $S_{M-1}^{\text{opt}}(D_2, ..., D_M, E_0 + E_1) = \sum_{t=2}^{M} N_t^{\text{opt}}(D_t, -d^{-1}(D_t))$. Thus, since $N_1^{\text{opt}}(D_1, E_1)$ is increasing in E_1 by Lemma 2, it is optimal to have $E_1 \leq \Big(\sum_{t=2}^{M} d^{-1}(D_t)\Big) - E_0$. Furthermore, since a feasible E_1 satisfies $E_1 \geq -d^{-1}(D_1)$ and $E_0 + E_1 \geq 0$, it must be lower bounded as $E_1 \geq \max(-E_0, -d^{-1}(D_1))$. This establishes that we may perform our optimization of E_1 over the interval $\mathcal{I}_{E_1}$. For a given E_1, the optimal photon number in the first time interval is given by $N_1^{\text{opt}}(D_1, E_1)$. The result of combining these observations is expressed in Equation (38).

In Figure 6, we show these results for three time intervals, fixing $E_0 = 1$, $\eta = 0.9$, and demands D_1, D_2, and D_3 summing to 5. Figure 6(a) shows the minimum average photon number plotted with respect to D_1 and D_2 (note that the lower right half of the figure is outside the region we are considering, where the three demands are nonnegative and sum to 5). Figure 6(b) displays the corresponding PIE. As we can see, the highest PIE is generally achieved when the demands are roughly equal, but it

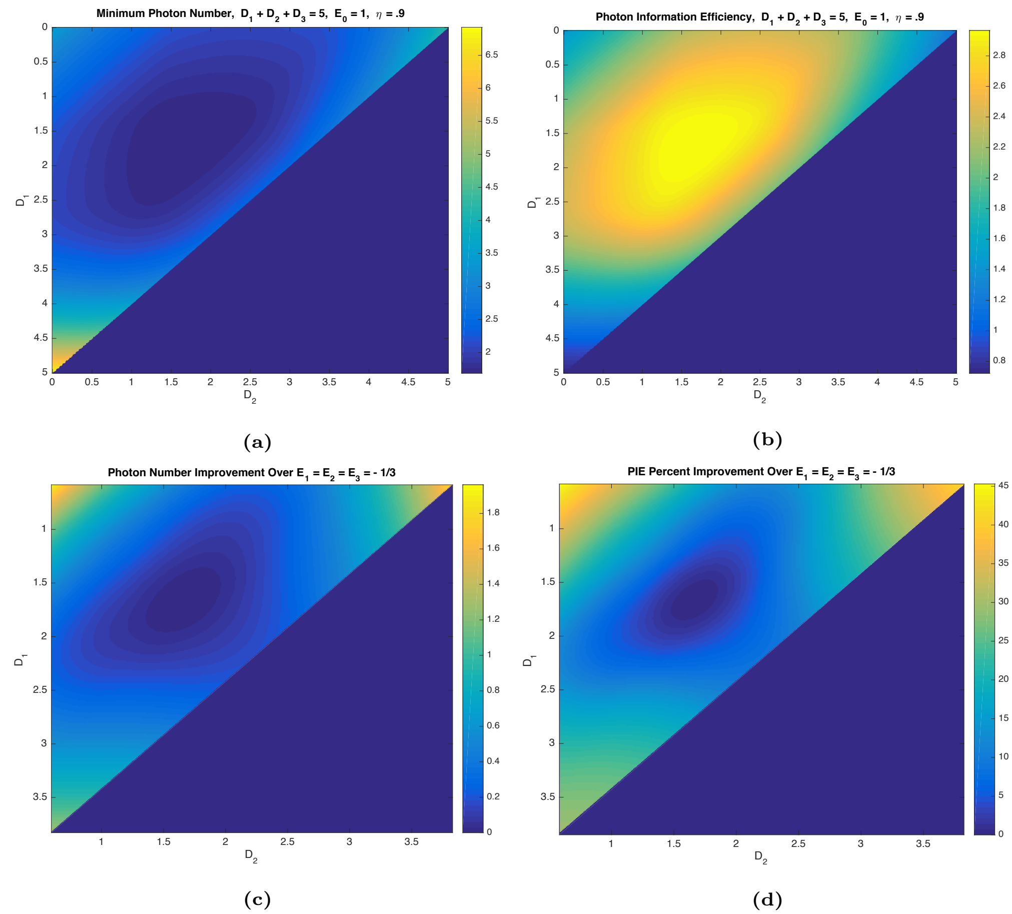


Figure 6. (a) Optimal average photon number for $M = 3$ time intervals with average bit demands D_1, D_2, and D_3 summing to 5, initial entanglement $E_0 = 1$, and transmissivity $\eta = 0.9$. (b) The associated photon information efficiency. (c) The reduction in average photon number over consuming the stored entanglement at a constant rate, $E_1 = E_2 = E_3 = -1/3$, plotted where optimization in Equation (17) is feasible for the associated values of D_1, D_2, and D_3. (d) The associated percent increase in PIE. All plots are with respect to (D_1, D_2), with the lower right half omitted since it falls outside the domain in which $D_1 + D_2 + D_3 = 5$.

is interesting to look at the three corners of the triangles formed in the figure, which correspond to when the bulk of the classical information is demanded in only one of the three time intervals. Among these three cases, the highest PIE corresponds to when $D_3 = 5$, in which case the first two time intervals can be spent generating extra entanglement to facilitate the classical communication. Conversely, the lowest PIE corresponds to when $D_1 = 5$, in which case the second two intervals are not used at all.

Figures 6(c) and 6(d) respectively show the corresponding reduction average photon number and percent increase in PIE compared to what we would achieve by communicating at a constant rate of entanglement consumption, $E_1 = E_2 = E_3 = -1/3$. Figures 6(c) and 6(d) are only plotted in the feasible region of optimization in Equation (17), where $-1/3 \geq -d^{-1}(D_t)$ for $t = 1, 2, 3$. As we can see, when the demands for classical information vary significantly over the three time intervals, optimizing our entanglement usage can significantly increase our PIE over a uniform rate of consumption, sometimes by as much as 45%.

VIII. Conclusion

We have demonstrated that over the pure-loss bosonic channel, using knowledge of time-varying demands for classical communication rates, we can optimize the rates at which we generate or consume entanglement, as well as the associated average photon number, in order to maximize our overall photon information efficiency. By formulating the problem as a finite-time discrete optimization, we were able to derive a concise description for the feasible region of rates of entanglement generation and consumption, as well as a functional form for the optimal average photon number at which to communicate at given rates of classical communication and entanglement usage. We used this to derive an expression for the solution to the entanglement battery problem for an arbitrary number M of time intervals, and to compute the optimal photon information efficiency for $M = 1$, 2, and 3. In particular, our results yielded cases in which we can benefit by using some of our photon resources to generate entanglement during times of relatively low demand for classical information, which we can then use to communicate at a lower average photon number during times of higher demand. Since we worked under the assumption that we can achieve the capacity region of (2) in each of our time intervals, some care must be taken to generalize our results to the case of a continuously time-varying demand profile. More work still is required to construct explicit codes and protocols which achieve the PIE values that we predict, and to extend our results to the thermal noise bosonic channel. We may also treat problems such as limits on the amount of quantum memory at our disposal with which to store entanglement, or limits on the amount of time over which we can reasonably store it. These limits manifest themselves naturally as new constraints on the entanglement variables E_t in our optimization. Another important avenue to pursue is to consider a nonzero demand for qubit communication ($Q > 0$) in our optimization. Arguably, one of the most important uses of entanglement is as a means for teleporting quantum states, which are much more difficult to communicate than classical bits. As such, an entanglement battery treatment of qubit communication would be very valuable. Our results represent an important step in designing advanced deep space optical communication protocols, allowing us to determine the average photon numbers and entanglement usage profiles required to maximize our overall photon information efficiency.

References

[1] C. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[2] S. Dolinar, K. M. Birnbaum, B. I. Erkmen, and B. Moison. On approaching the ultimate limits of photon-efficient and bandwidth-efficient optical communication. *IEEE*

International Conference on Space Optical Systems and Applications, pages 269–278, 2011.

[3] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen. Classical capacity of the lossy bosonic channel: The exact solution. *Physical Review Letters*, 92(2):027902, 2004.

[4] A. S. Holevo. The capacity of the quantum channel with several general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, January 1998.

[5] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, July 1997.

[6] S. Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55(3):1613–1622, March 1997.

[7] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51(1):44–55, January 2005.

[8] P. W. Shor. The quantum channel capacity and coherent information. *MSRI Workshop on Quantum Computation Lecture Notes*, 2002.

[9] I. Devetak and P. W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Communications in Mathematical Physics*, 256(2):287–303, June 2005.

[10] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895–1899, March 1993.

[11] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69(20):2881–2884, November 1992.

[12] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Physical Review Letters*, 83(15):3081–3084, October 1999.

[13] P. W. Shor. *Quantum Information, Statistics, Probability (Dedicated to A. S. Holevo on the occasion of his 60th Birthday): The classical capacity achievable by a quantum channel assisted by limited entanglement*. Rinton Press, Inc., 2004.

[14] M.-H. Hsieh and M. M. Wilde. Entanglement-assisted communication of classical and quantum information. *IEEE Transactions on Information Theory*, 56(9):4682–4704, September 2010.

[15] M.-H. Hsieh and M. M. Wilde. Trading classical communication, quantum communication, and entanglement in quantum Shannon theory. *IEEE Transactions on Information Theory*, 56(9):4705–4730, September 2010.

[16] M. M. Wilde, P. Hayden, and S. Guha. Quantum trade-off coding for bosonic communication. *Physical Review A*, 86:062306, 2012.

[17] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, and J. H. Shapiro. Minimum output entropy of bosonic channels: A conjecture. *Physical Review A*, 70:032315, 2004.

[18] M. M. Wilde. *Quantum Information Theory, 2nd edition*. Cambridge University Press, 2016.

[19] M. M. Wilde, P. Hayden, and S. Guha. Information trade-offs for optical communication. *Physical Review Letters*, 108(14):140501, 2012.